Modelos mixtos y otras extensiones

Javier Fernández-López, Profesor Ayudante Doctor
Unidad de Matemática Aplicada
Departamento de Biodiversidad, Ecología y Evolución, UCM

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

DOKTOREGO ESKOLA
ESCUELA DE DOCTORADO
DOCTORAL SCHOOL

Programas de doctorado:

- Agrobiología Ambiental
- Calidad y Seguridad Alimentaria
- Biodiversidad, Funcionamiento y Gestión de Ecosistemas

Bloque 3: Modelos mixtos y otras extensiones

- Factores de efectos aleatorios VS factores de efectos fijos
- Efectos aleatorios: intercepto y pendiente (interacciones)
- Concepto de autocorrelación
- Otros controles de la no-dependencia: autocorrelación temporal y espacial

EFECTOS FIJOS VS EFECTOS ALEATORIOS

- Hasta ahora hemos aprendido a utilizar predictores categóricos en los cuales **estábamos interesados**: sexo, tipo de hábitat, etc. Estos eran tratados como **efectos fijos**.

- Por el contrario, los **efectos aleatorios** son parámetros estadísticos que intentan explicar el **ruido** causado por agrupamientos o *clústeres* en los datos que intentamos modelizar.

EFECTOS FIJOS VS EFECTOS ALEATORIOS

- Hasta ahora hemos aprendido a utilizar predictores categóricos en los cuales **estábamos interesados**: sexo, tipo de hábitat, etc. Estos eran tratados como **efectos fijos**.

- Por el contrario, los **efectos aleatorios** son parámetros estadísticos que intentan explicar el **ruido** causado por agrupamientos o *clústeres* en los datos que intentamos modelizar.

- Estudiantes en diferentes colegios. Niveles del factor colegio?
- Individuos en diferentes poblaciones. Niveles del factor población?
- Muestras en diferentes establecimientos. . Niveles del factor establecmiento?
- Hojas en diferentes árboles. Niveles del factor árbol?

EFECTOS FIJOS VS EFECTOS ALEATORIOS

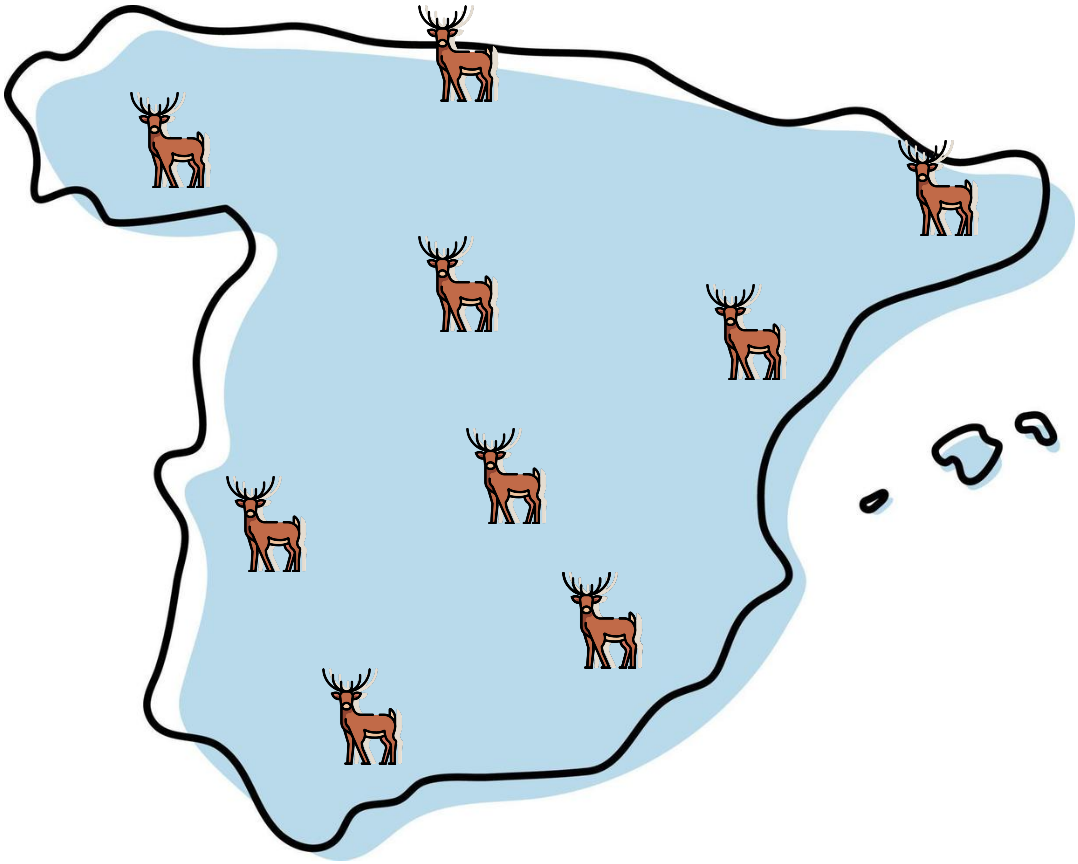
- Hasta ahora hemos aprendido a utilizar predictores categóricos en los cuales **estábamos interesados**: sexo, tipo de hábitat, etc. Estos eran tratados como **efectos fijos**.

- Por el contrario, los **efectos aleatorios** son parámetros estadísticos que intentan explicar el **ruido** causado por agrupamientos o *clústeres* en los datos que intentamos modelizar.

- Estudiantes en diferentes colegios. Niveles del factor colegio?
- Individuos en diferentes poblaciones. Niveles del factor población?
- Muestras en diferentes establecimientos. . Niveles del factor establecmiento?
- Hojas en diferentes árboles. Niveles del factor árbol?

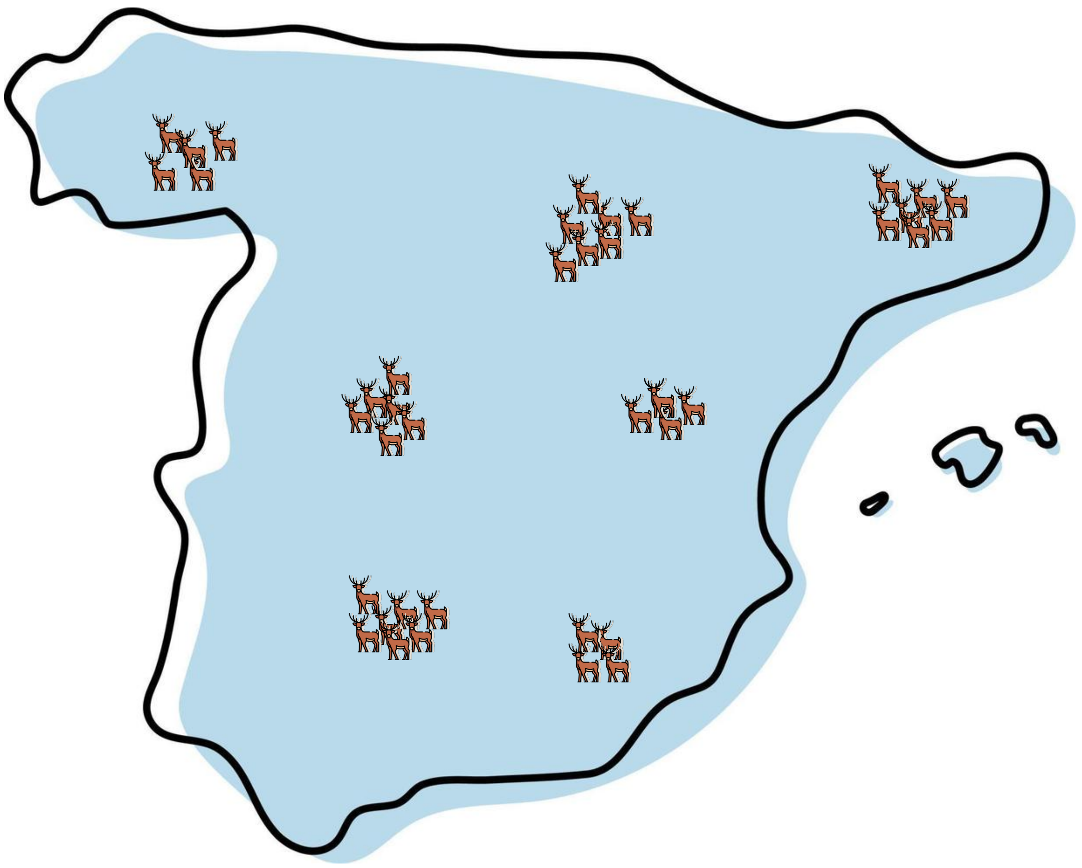
- Sexo o clases de edad no son aleatorios: normalmente estamos interesados en su efecto, y si volviésemos a muestrear, encontraríamos los mismos niveles para estos factores.

EFECTOS FIJOS VS EFECTOS ALEATORIOS



¿Cómo afecta la temperatura a la abundancia de garrapatas en ciervo?

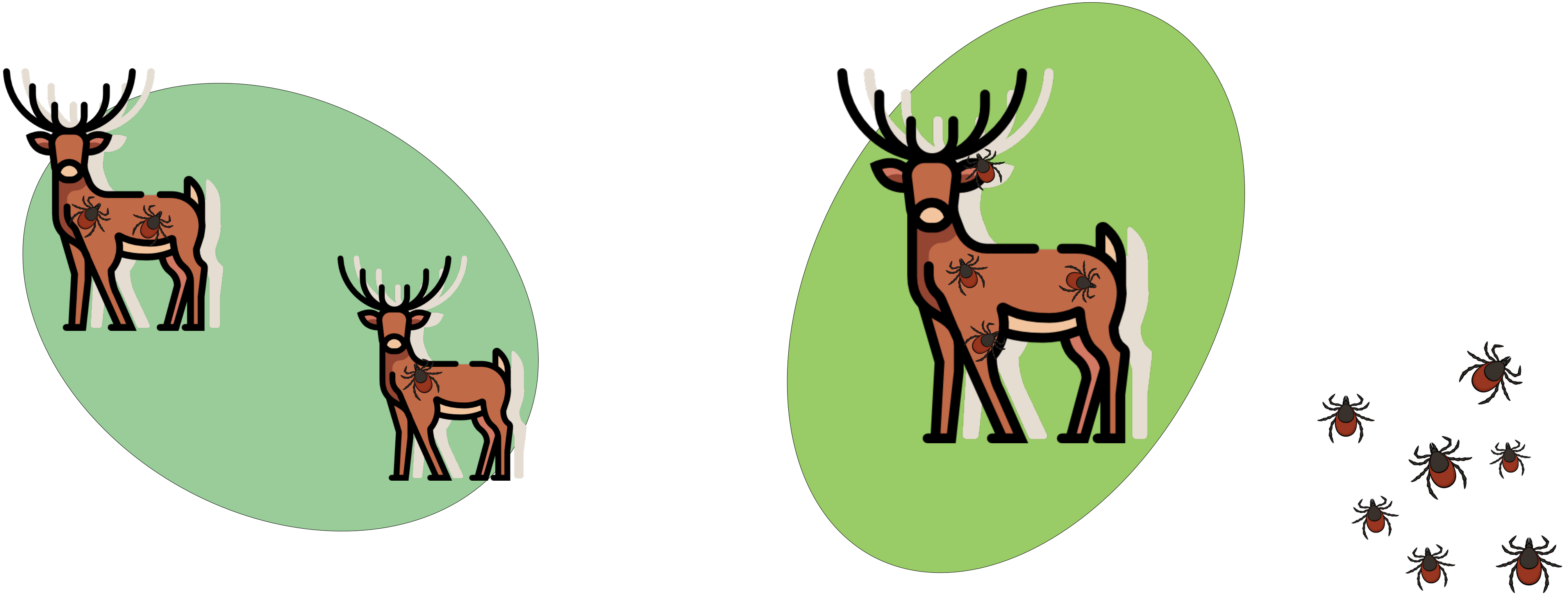
EFECTOS FIJOS VS EFECTOS ALEATORIOS



¿Cómo afecta la temperatura a la abundancia de garrapatas en ciervo?

EFECTOS FIJOS VS EFECTOS ALEATORIOS

¿Cómo afecta la temperatura a la abundancia de garrapatas en ciervo?



Diferencias entre efectos fijos y efectos aleatorios

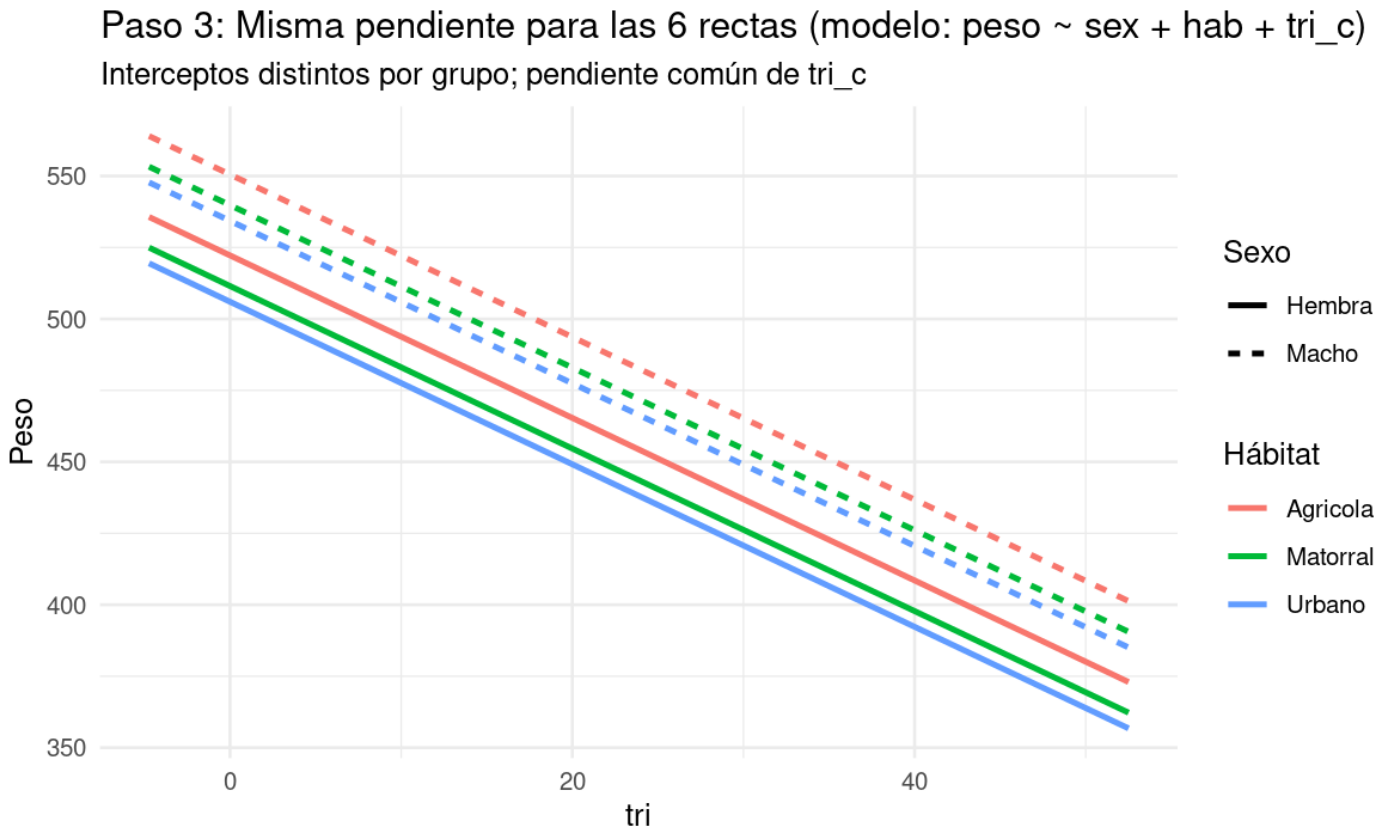
Factores de efectos fijos: variables de **interés** específico. Suelen tener un "numero limitado" de niveles. Si volviésemos a obtener una muestra, obtendríamos los mismos niveles.

Factores de efectos aleatorios: capturan la variabilidad de factores que queremos **controlar**, pero que en realidad no estamos interesados en ellos (efectos atribuidos a ser una muestra de una población mayor). Suelen tener "muchos niveles". Si volviésemos a obtener una muestra, podríamos obtener otros niveles.

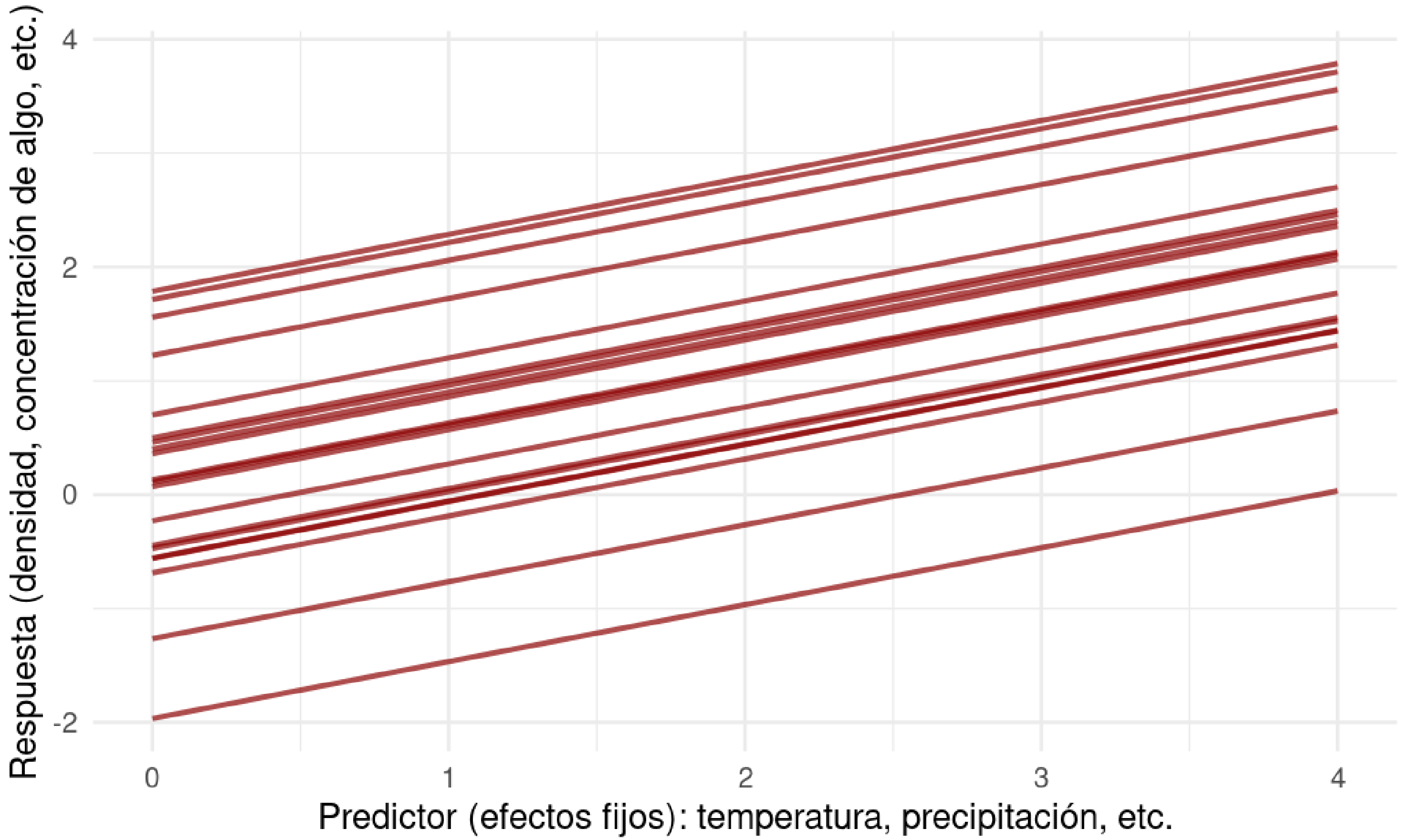
MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO

$$\mathrm{peso}_i \sim \mathrm{Normal}(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \beta_1\, I(\mathrm{sex}_i = \mathrm{Macho}) + \beta_2\, I(\mathrm{hab}_i = \mathrm{Matorral}) + \beta_3\, I(\mathrm{hab}_i = \mathrm{Urbano}) + \beta_4\, \mathrm{tri}_i$$



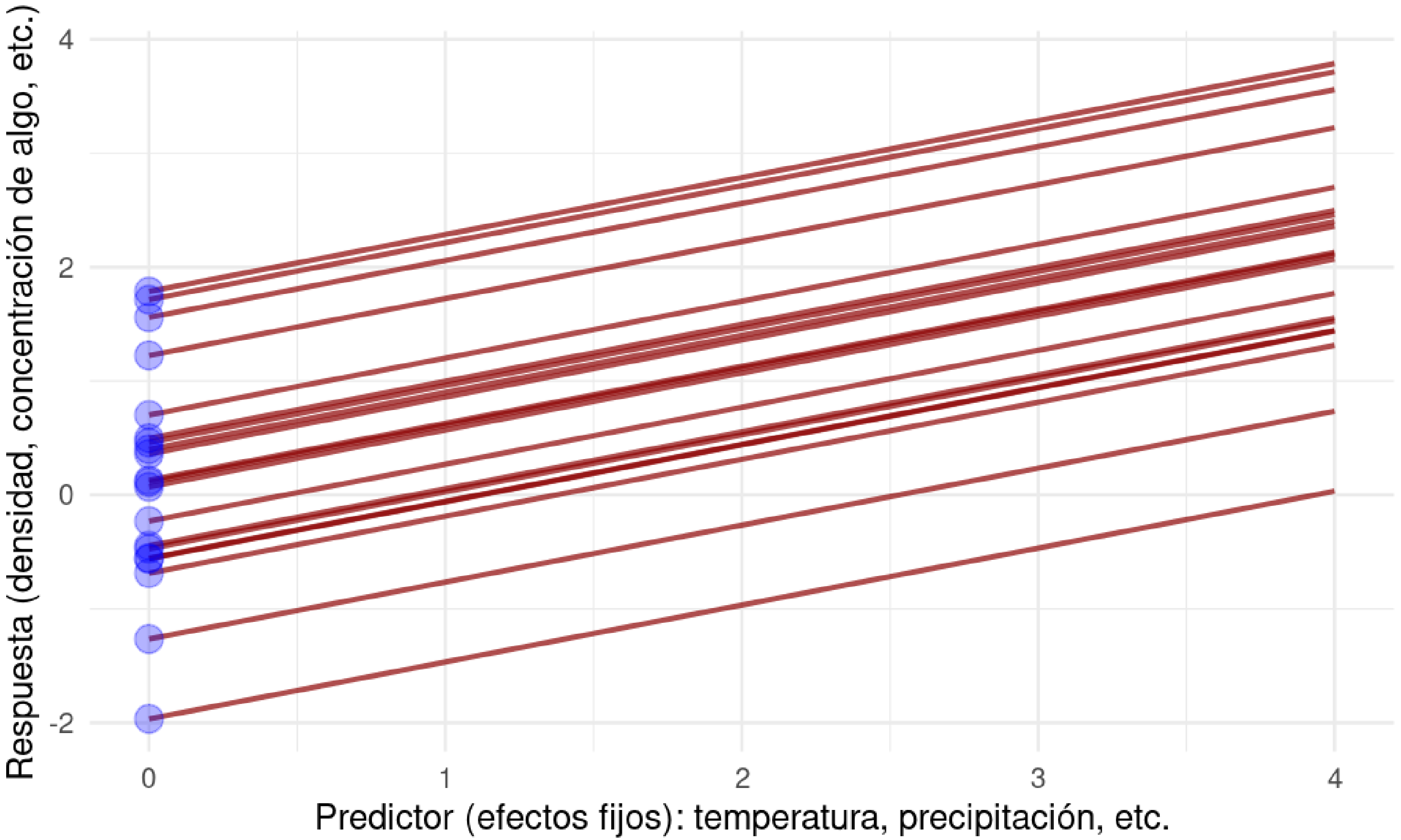
MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO



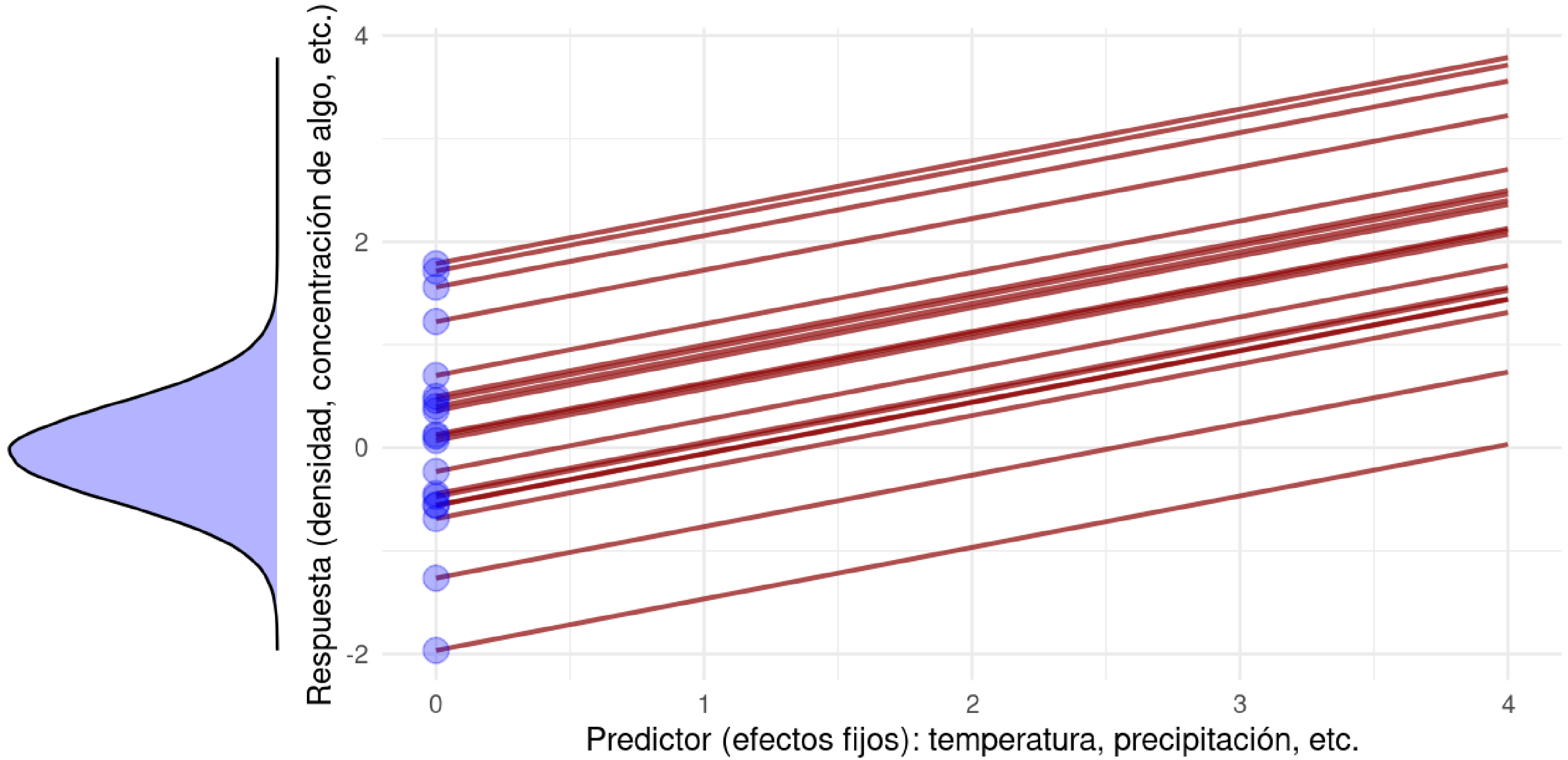
MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO

Intercepto de efectos fijos

Valores estimados uno a uno



MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO

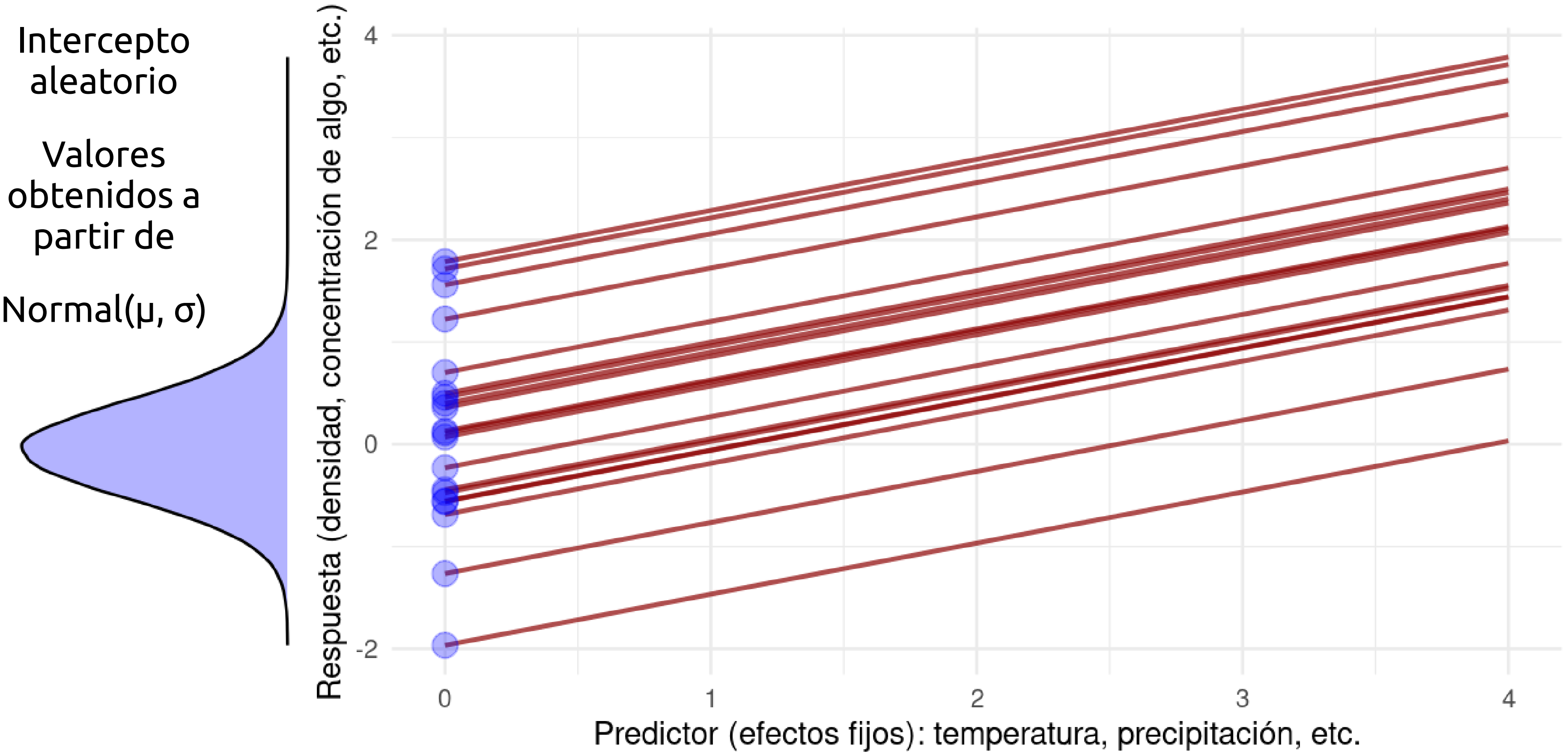


MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO

Intercepto aleatorio

Valores obtenidos a partir de

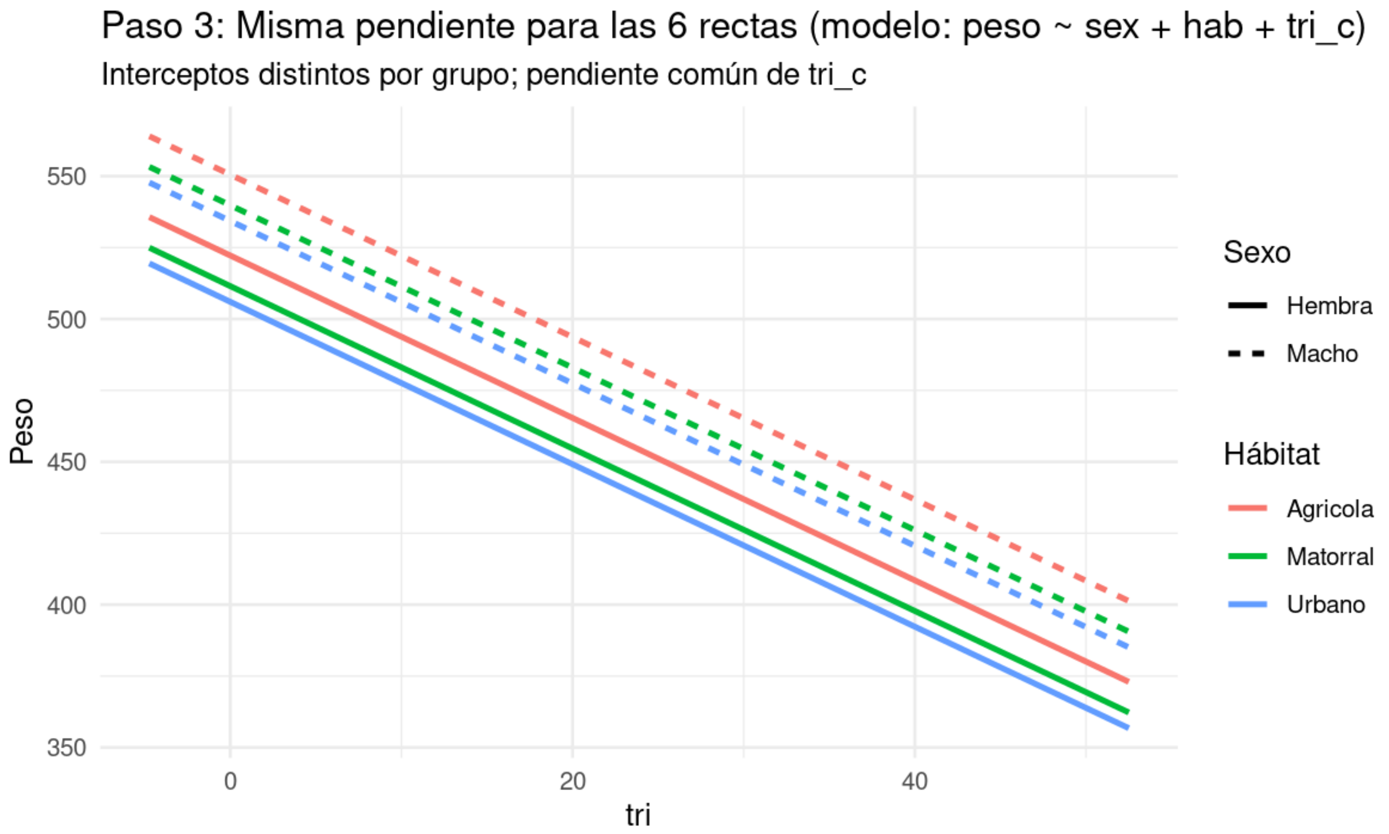
Normal(μ, σ)



MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO

$$\text{peso}_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \beta_1\, I(\text{sex}_i = \text{Macho}) + \beta_2\, I(\text{hab}_i = \text{Matorral}) + \beta_3\, I(\text{hab}_i = \text{Urbano}) + \beta_4\, \text{tri}_i$$



MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO

$$\text{nº garrapatas}_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1\,\text{temperatura}_i + \beta_3\,\text{cotoB}_i + \beta_4\,\text{cotoC}_i + \;\cdots\; + \beta_{28}\,\text{cotoZ}_i$$

Si usásemos el coto como un factor de efectos fijos, nuestro modelo tendría que calcular un parámetro para cada coto… por lo que necesitaríamos bastantes ciervos dentro de cada coto para tener estimas robustas.

Además, se podría decir que "se pierde información"… estáis seguros de que el número de garrapatas que encontramos en el coto A no se va a parecer en nada a las del coto B, a las del coto C…? En palabras de Richard McElreath, estos modelos "*tienen amnesia*"!

MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO

$$\text{n}^{\text{o}}\ \text{garrapatas}_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \beta_{\text{coto}(j)} + \beta_1\ \text{temperatura}_i$$

$$\beta_{\text{coto}(j)} \sim \mathcal{N}(0,\ \sigma_{\text{coto}})$$

Podemos usar el coto como un factor aleatorio. Esto quiere decir que asumimos que todos esos interceptos provienen de una misma distribución, una normal con media 0 y desviación estándar que es el parámetro que tenemos que estimar. Esto simplifica el modelo, a costa de asumir que esos interceptos provienen de la misma distribución, es decir, el modelo asume que **se comparte información** entre los cotos. Estos modelos "*tienen memoria*"!

El número **efectivo** de parámetros no es exacto en este tipo de modelos, sino que varía entre 2 y 29.

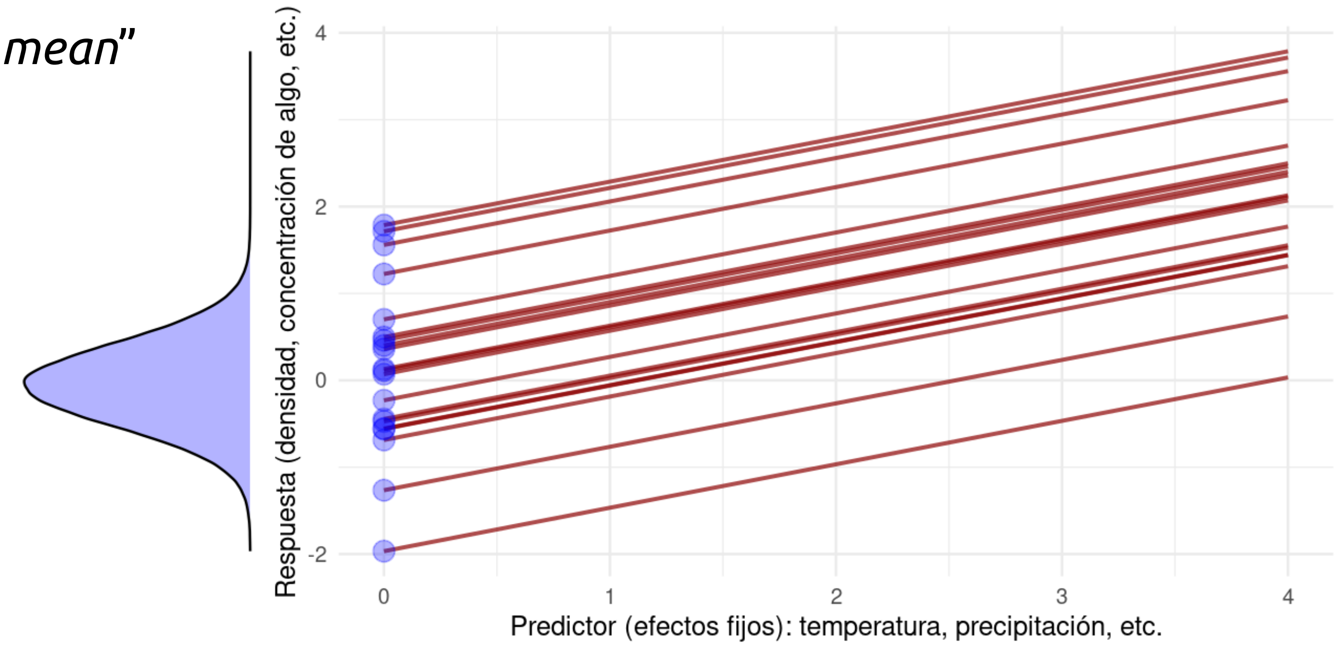
MODELOS MIXTOS: EL ÁLGEBRA DEL MODELO

$$\text{n}^{\text{o}}\ \text{garrapatas}_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \beta_{\text{coto}(j)} + \beta_1\ \text{temperatura}_i$$

$$\beta_{\text{coto}(j)} \sim \mathcal{N}(0,\ \sigma_{\text{coto}})$$

El efecto "*shrinkage to the mean*"



BONDAD DE AJUSTE EN GLM

- En este tipo de modelos, es muy común que la mayor parte de la variación de nuestros datos venga explicado por los efectos aleatorios, sobre todo cuando trabajamos con datos fuera del laboratorio con mucho "ruido".

- Por eso, es común diferenciar entre la proporción de la varianza explicada por los efectos fijos (en los cuales estamos realmente interesados, la temperatura en nuestro caso, **R^2 marginal**), de la proporción total de varianza explicada por el modelo (efectos fijos + efectos aleatorios, **R^2 condicional**)

- El R^2 de **Nakagawa** sirve a este propósito.